

Abnormal patient, normal test: the otologist's dilemma

A daily and frustrating occurrence in clinical otology is the patient with a variety of distressing complaints despite normal or non-diagnostic tests. What does this mean, and how should we deal with the problem?

Medical diagnoses are generally made based on history and physical examination, followed by the appropriate tests. In addition to blood and urine testing, we look specifically at structure and function. The patient presents with a number of complaints, and we try to select tests which will confirm our clinical impression, rule out other items in the differential diagnosis, and pick up unexpected findings.

In clinical otology, standard tests of function are a rather meagre handful: audiometric testing, auditory brainstem response (ABR) testing, electrocochleography and vestibular evaluation. The test–patient discrepancy has numerous causes and clinical ramifications. Let's look at just a few.

Beginning with the humble tuning fork, consider the Rinne test. For the 512 Hz fork, the Rinne test turns negative (i.e. bone conduction is better than air conduction) between 30 and 45 dB conductive hearing loss. A simple screening test, but definitive? It will miss a 25 dB conductive loss. For that, one needs a 256 Hz fork, for which the Rinne test will pick up conductive losses between 15 and 30 dB. While using sets of tuning forks (from 128 to 1024 Hz) to quantitate hearing loss was common practice in the past, today the basic evaluation of a hearing impaired patient requires audiometry.

So, let's look at the audiogram, the workhorse of clinical otology. This represents a psychoacoustic dialogue between patient and audiologist, with many inaccuracies and confounders. It is well known in quantum physics that even the most passive observer alters a physical phenomenon, simply by observing it. How much more, then, is this the case in an audiometric test? Test–retest variability tells only part of this story.

In pure tone testing, consider that we are sampling at octave frequencies (and using natural octaves, not the tempered octaves of the piano keyboard, which is why the beeps sound flat). Even if we think of pitch as progressing just in semitones (and in reality the ear can distinguish several finer gradations in frequency),

there are 12 steps from 1 octave to the next. This means that 11 pitches are not tested. Although we occasionally test fifths rather than octaves (e.g. for 3 kHz), for the most part the pure tone test is really a very coarse screen of hearing acuity, a 'Swiss cheese' screen with more holes than substance. A patient may well have a significant hearing loss that 'falls between the cracks' of standard pure tone testing. (In contrast, the Bekeby test, which assessed all frequencies equally, was a much more comprehensive assessment of hearing; of course, such testing took longer and was more labour-intensive.)

So, the audiogram misses more frequencies than it catches. But is a slight hearing loss, at say 2354 Hz, really that important, or is it just an irrelevant piece of data? Setting aside the instinctive answer, which is that the more we know the more we can understand, consider the patient who complains of tinnitus in the face of a normal standard audiogram, or the patient with impaired sound discrimination because some specific frequencies in the consonant range are not heard, even though hearing at 2, 4 and 8 kHz is 'normal'. Does this patient need ABR testing, or just a more diagnostic pure tone test? The cochlea can tell us much more than what we plot as we connect the dots on a standard audiogram.

Moving to speech discrimination, the testing paradigm becomes critically interactive. What is presented is affected by the audiologist's voice, by his or her vocal pitch, loudness and accentuation, and also by his or her finesse, persistence and time constraints. The patient's level of attention, co-operation and comprehension are additional variables.

The statistics of speech scoring are also deceptive. A speech recognition test result of 60 per cent does not mean that the patient understood 60 out of 100 words. The usual short cut here is a list of 25 words, with each correct answer adding 4 per cent, but in some practices this has been pared down to only 20 words. Now, if each word represents 5 per cent, is a 10 per cent change in discrimination really significant? How about a 15 per cent change? Does the speech reception threshold (SRT) result really mandate ABR testing or magnetic resonance imaging?

While the deficiencies of the audiogram hide ‘in plain sight’, the limitations of electrocochleography and ABR testing are generally recognised, and, unlike audiography, most clinicians use these latter tests as guides rather than sources of definitive diagnostic data. Such tests indicate trends rather than provide answers – trends which become significant if they confirm clinical findings, but which can confound if not. This is especially true of electrocochleography, the results of which can vary from audiologist to audiologist, from day to day and from test to test. In this regard, it would be interesting to design a ‘halter electrocochleogram’ which monitored the electrical responses of the cochlea over longer periods of time. Interesting normative data relating to barometric pressure, diet and hormonal influences might emerge.

What about vestibular testing? Unlike audiometry, caloric testing is objective. Since Barany’s pioneering work, this test has been basically unchanged, and apart from minor ‘tweaks’ in technique (using air instead of water, for example), the parameters have been fixed, the results standardised and reproducible. Surely caloric testing has stood the test of time?

But is it a good test? The eminent auditory scientist Juergen Tonndorf once said that testing vestibular function with calorics is like testing vision with lightning and hearing with thunder. Caloric testing is not physiological, but a massively supra-threshold stimulus that does not address the labyrinth in a language it normally understands. The meaning of this is simple but profound: a patient may have a clinically significant vestibular weakness with a statistically normal caloric test. And, while there are more physiologically appropriate tests of semicircular canal function, such as the rotary chair, the front-line reality remains that caloric tests are easy, cheap, and the standard in clinical practice.

So it seems that our otological tests miss many abnormalities, for several reasons, including: the size of the holes in our screening techniques; the appropriateness of the set-point for what is considered significant; and the technical problems inherent in making tests reliable and consistent. Otological tests are particularly weak at capturing momentary abnormalities. Tests are snapshots, and life is a movie: unless the test is performed at the moment a transient phenomenon occurs, the report may well come back as ‘normal’. This is particularly a problem in cases of transient dizziness or tinnitus.

There is however a deeper, philosophical issue to consider. What is the definition of ‘significant’? For otological tests, such a definition is circular. We state that caloric testing identifies significant vestibular weakness, then define significant vestibular weakness as one that causes an abnormal caloric response. Instead of being self-referential, these tests need to be patient-referential. ‘Significant’ needs to be defined as a sign that something is wrong with the patient, not with the test. As the Greeks said, ‘man is the measure of all things’. The graphs and numbers

generated in otological testing serve a function only if they are measured against the patient’s problems.

Physicians live in a visual world, and we are willingly seduced by pictures, numbers, figures and graphs. Visual data generate an apparently irrefutable point of reference for patients, health workers, health insurance companies and lawyers, especially when such data validate our clinical impression. Visual data are static and easily shared to build consensus among their interpreters. And so, clinicians are increasingly becoming testers: record generators rather than clinical observers. However, our reliance on (some would say addiction to) clinical testing may lead us in a false direction that does not benefit our patients. The test is just a tool, it is not the disease. As the Buddhist injunction warns, ‘do not confuse the finger pointing at the moon, with the moon’. So, while a caloric weakness of 15 per cent may not in itself be ‘statistically significant’, when put into the context of a dizzy patient who lateralises to the same side, it becomes ‘the finger pointing to the moon’.

We have now questioned the tests, and it seems that they have many limitations: they can miss and, worse, mislead. But we still need to deal with that abnormal patient with normal tests. What to do? Trying to convince the patient that ‘your tests are normal and nothing is wrong’ is disingenuous, and just as inadequate as the defensive posture of ‘the patient must be crazy’.

So which is correct, the patient or the test? We need to accept that actuality trumps theory. As a young medical student in Toronto in the 1960s, I recall performing cadaver dissection and finding something that was not in the anatomy manual. Confused, I sought out the demonstrator, an old English doctor with years of experience. ‘Sir’, I said anxiously, pointing to the structure in question, and then to the book, ‘this is not supposed to be here!’. He looked at me calmly, and uttered words that have continued to resonate ever since: ‘My boy, never argue with the specimen’.

Patient trumps test. We need to soberly accept the reality that these tests are rather blunt instruments which may signify but do not conclude. In otology, there are many symptoms and few tests. We have no way to quantify diplacusis, hyperacusis or pain accompanying inner ear symptoms, to give just a few examples; however, that doesn’t mean that these complaints are any less valid, and no amount of ‘normal testing’ will make them disappear. Patients will often convey to us symptoms which do not fit our mental picture of a ‘diagnosis’, and which are not confirmed by our tests. We should not discard these complaints simply because they do not fit. What is, is, and whether or not such symptoms fit the crude parameters imposed by tests does not make them any less real. And making the patient fit the diagnosis, forcing him or her to lie down on the Procrustean bed of our tests and theories, serves neither patient nor physician. (Greek

mythology tells of the highwayman Procrustes who invited travellers to his house and offered them overnight lodgings. Once they accepted, however, he made them lie down on his iron bed and, being a stickler for exactitude, would stretch them on a rack if they were too short or cut off their legs if they were too long, until they fit the bed exactly. Some of his victims died in the adjustment.)

So if the patient's clinical picture and our tests don't match, we should not question only the patient, but the tests too. Clinical tests are not an

end, merely a means. They are tools: imperfect, rough, and helpful only if used with a clear understanding of their limitations, and always with the clinical picture foremost in our mind. Remember, it's the patient, not the test.

ANTHONY F JAHN
Professor of Clinical Otolaryngology
Columbia University College of Physicians and
Surgeons
New York, USA